DOCUMENT RESUME

ED 225 356 FL 013 397

AUTHOR Mohr, Eugene V.
TITLE Examining the English Examination. A Study of the Linguistic Content of ESLAT.
INSTITUTION College Entrance Examination Board, Hato Rey, PR. Puerto Rico Office.
PUB DATE 78
NOTE 20p.
AVAILABLE FROM Puerto Rico Office of the College Entrance Examination Board, Apartado 1275, Hato Rey, Puerto Rico 00919 ($3.00).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *English (Second Language); Error Patterns; *Interference (Language); *Language Tests; Secondary Education; Standardized Tests; *Stress Variables; Testing; *Test Validity
IDENTIFIERS *College Entrance Examination Board; Puerto Rico

ABSTRACT

An analysis is presented by the Puerto Rico office of the College Board of the English as a Second Language Achievement Test (ESLAT), which was undertaken to measure the specifications and item content of the examination against the students' performance. The report, based on 400 randomly selected structure items used in one or more versions of ESLAT, focuses on an examination of differences in question form, success of the distractors, linguistic factors affecting item difficulty, and ways these factors operate on student responses to certain areas of grammar. The linguistic factors examined are frequency of occurrence in common usage, range of applicability, variability, interference, and reinforcement of an English language pattern by a native language pattern. After the study of the interplay of these factors, the language areas tested by the 400 items are divided into four levels of difficulty and significant correlations between certain language areas and difficulty levels are discussed. In addition to the structure items, 10 reading selections are examined and discussed. Finally, an attempt is made to relate the ESLAT results to the emphases in the textbooks used in Puerto Rican public high schools. It is concluded that, while the ESLAT is an excellent testing instrument, interference problems are still a major source of difficulty for English as a second language students. (AMH)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED225356

# EXAMINING THE ENGLISH EXAMINATION

## A Study of the Linguistic Content of ESLAT

By Eugene V. Mohr

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jorge J. Deppa
CEEB/Puerto Rico

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL013 397

COLLEGE ENTRANCE EXAMINATION BOARD

Copies of this book may be ordered from the Puerto Rico Office of the College Entrance Examination Board, Apartado 1275, Hato Rey, Puerto Rico 00919. The price is $3.00.

*Copyright © 1978 by the College Entrance Examination Board.*
*All rights reserved. Derechos reservados.*
*Printed in the United States of America.*

**About the Author**

Dr. Eugene V. Mohr has lived and taught in Puerto Rico for close to twenty-five years. Most of that time has been dedicated to teaching courses in English grammar, linguistics and literature, directing the departments of English at the San Juan campus of Inter American University and the Río Piedras campus of the University of Puerto Rico, and serving as language consultant for numerous public and private organizations. A respected member of the local educational community, Dr. Mohr has been Chairman of the Committee of Examiners of the English as a Second Language Achievement Test and the Advanced Level Examination in English, both developed by this Office. He also served as a member of the Commission to Study the Teaching of English in the High Schools of Puerto Rico, a joint effort of the College Board and the Puerto Rico Department of Education.

# Foreword

In keeping with a College Board tradition of encouraging continuous public discussion of its programs, Professor Eugene V. Mohr was asked to make an independent and critical review of the English as a Second Language Achievement Test (ESLAT). The latter is part of the Board's admission battery for post-secondary institutions in Puerto Rico and is constructed with the counsel and cooperation of advisory committees composed of teachers and administrators from local schools and colleges.

The assignment to Professor Mohr was prompted by the permanent interest of the Board in assuring that ESLAT possess—apart from the consideration of its validity and reliability in predicting college performance—the desirable pedagogical properties and functions that should characterize each and every one of the Board's tests. In this particular instance, the principal objective was to measure the specifications and item content of ESLAT against the students' performance in order to study the test items and procedures that best lend themselves to a descriptive and qualitative analysis of English language attainment in Puerto Rico. Such a revision as this inevitably examines the curricular validity of test items, passes judgment on their suitability and provides an effective check against "item obsolescence." It may also point to specific achievements as well as to areas of particular learning difficulties. And hopefully, its feedback effect on educational leadership will help generate concrete improvements in the planning of the English curriculum.

Studies such as this, although useful, are not in themselves a sufficient measurement of the complexities of teaching, in this case, English as a second language. Tests are by definition limited instruments and as such they should never become the sole rationale for judging the dynamic processes, formal and informal, involved in the teaching and learning of any subject matter. The intricate relationships intersecting tests, education and society will never allow for that type of simplification. Yet, the careful study of performances by thousands of students as they are recorded and statistically analyzed in the Board's testing programs is certainly a responsible and reliable way to improve the intellectual and academic descriptions of teaching and learning processes.

Professor Mohr addresses his findings and recommendations not only to the Board, but to the educational community in general. All of his suggestions have been well received by the College Board's office in Puerto Rico, and a good number will be implemented in the development of future tests. Others will be brought before the consideration of the proper advisory committees. As was implicit in the original commitment with Professor Mohr, this report was intended for publication and for distribution among the interested members of our educational community. We expect that it will be widely discussed, especially among teachers of English at the high school and post-secondary levels. Along these lines, particular attention should be given to the suggestion made in the report that the College Board and the Department of Education of Puerto Rico sponsor a joint project "to use ESLAT as a device for evaluating the relative effectiveness of different sets of teaching materials."

As a result of our conviction that this report is a significant contribution to educational evaluation in Puerto Rico, plans are being made for similar studies based on the Board's achievement tests in Spanish and mathematics. We believe that this should be done soon, and we are in the process of identifying knowledgeable and concerned professionals for the purpose.

Finally, the Puerto Rico office of the College Board expresses its appreciation of Professor Mohr's intellectual contribution and recognizes the public spiritedness implicit in his acceptance of such an assignment.

Adolfo Fortier-Ortiz
May 1978

# Part A: An Analysis of 400 Structure Items

1. *Goals and Methods*

The English as a Second Language Achievement Test (ESLAT) is designed to measure the English language proficiency of students graduating from high schools in Puerto Rico and planning to enter local colleges and universities. It is constructed by the Puerto Rico Office of The College Board with the cooperation of a broad spectrum of the educational community. Over the years the test has proven itself a reliable measuring instrument for overall proficiency in English.

The primary aim of this report, undertaken at the invitation of the College Board, is to point out specific English language structures which students find particularly difficult, so that teachers and school administrators can have this information available when adjusting the emphases in their English programs to improve student performance. Another aim is to suggest improvements that might be made in future versions of ESLAT through changes in the types and proportions of items included. All recommendations are accompanied by linguistic analyses of the problems under consideration. In addition, the textbooks used in Puerto Rico's public high schools were examined for possible correlations between teaching materials and ESLAT results.

The first seven sections of this report are based primarily upon 400 randomly selected structure items used in one or more versions of ESLAT. The items were arranged in four equal groups of ascending difficulty; the norm of difficulty used was the items' equated delta score, the easiest item examined having a delta of 6.5 and the most difficult item a delta of 19.7. The delta ranges of the four groups were the following:

| Group | I | II | III | IV |
|---|---|---|---|---|
| Delta range | 6.5 - 11.7 | 12.0 - 14.5 | 14.9 - 15.7 | 15.8 - 19.7 |

The delta is determined by the number of students who chose the correct response (the key) to an item as compared to the number who either chose one of the four incorrect responses (the distracters) or omitted the item. Some idea of the significance of the delta score can be formed from the following data on five items of varying difficulty.

| Delta | Correct Responses | Incorrect Responses | Omitted |
|---|---|---|---|
| 6.5 | 1829 | 165 | 1 |
| 9.4 | 1566 | 431 | 3 |
| 12.1 | 735 | 380 | 10 |
| 15.0 | 111 | 138 | 4 |
| 17.2 | 77 | 229 | 1 |
| 19.7 | 115 | 539 | 0 |

The delta varies slightly each time an item is used; for items which had been used more than once, the last recorded delta was the one selected.

Since many of the items studied are still in the active file, complete items are cited in the report only when absolutely necessary, when the distracters as well as the key are necessary to illustrate a point. At other times only the lead and key are cited, with the key underlined. For example, if the lead is *The chicken is* _____ *the pot*, followed by a choice of five prepositions, the item is cited as *The chicken is in the pot*. Finally, some of the illustrations are not real items at all, but simply made up illustrations.

As might be expected, the clearest patterns below emerged from contrasts between Group I and IV, the easiest and most difficult items. Nevertheless, it was gratifying to find areas where the patterning continued uniformly through all four groups. These were the patterns which yielded most information.

It should be recognized from the start that a study of this sort had certain built-in limitations. It was the first such study made of ESLAT, so it had no model. It was based on materials designed with different objectives in mind. It lacked the support of planned, controlled experiments. On the positive side, it derived from a unique opportunity—that of using the ESLAT files—and was guided by intuitions drawn from many years of experience in the fields of English as a second language (ESL) and Spanish-English contrastive analysis. Hopefully it will be supported by the intuitions of others with similar experience and will act as a catalyst to further research and experimentation.

2. *Form Differences*

The linguistic content of an item—the feature being tested—is contained in the lead and key. The same content may be tested in different ways, and such differences are *form* as opposed to *content* differences. Form differences examined in the ESLAT items included shapes and relationships of lead and key, apparent probability of distracters, failure to distinguish between phonemic and graphemic patterns, length of items and type of response (correct or incorrect) required.

The most common form of item was the fill-in type, in which the lead consisted of a sentence, part of which was left blank to be completed by the key. Another common form consisted of five "sentences," with instructions to choose the correct or most acceptable one. A third form required the student to correlate a question-word question with an appropriate answer or to select an appropriate question word to match a given response (*_____ are you going tomorrow? To San Juan*). This group is distinguished from other fill-in items because it involves a relationship between two structures. Finally, echo-verb items test the ability to repeat the essentials of a given verb structure by selection of the correct tense and auxiliary in a negative or affirmative context; there are three basic types of echo-verb items:

> He left early, *didn't he*?
> He has been paid, but I *haven't*.
> "You're angry." "No, *I'm not*."

Again, echo-verb items are listed as a separate form because of their distinct structural characteristics and because of a distinguishing pattern of difficulty among them.

The most striking correlation between form and level of difficulty was found among the choose-the-correct-sentence items. Group I contained 28 items of this sort, whereas Group IV had only seven. No conclusion can be drawn from this fact, however, until another correlation, between form and content, is taken into account: 27 of the 28 correct-sentence items in Group I tested word order. Word-order items accounted, in fact, for a third of the items in Group I and they decreased by about a half at each succeeding level of difficulty:

Group I: 33 WO items  Group II: 15  Group IIII: 7  Group IV: 4

Why were these correct-sentence word-order items so easy? Probably for reasons of both form and content.

In terms of content, one reason is that, except for a few easily understood differences, the order of major sentence units is very similar in Spanish and English. Also, since word order occurs in all except single-word sentences, it is practiced more often than any other grammatical feature.

There are also good reasons why the form of these items would make them relatively easy. First, they ask for the selection of a complete sentence pattern, the type of pattern most commonly practiced by students in ESL programs; it is assumed that students would respond more accurately to materials tested in the same format in which they were learned. Second, because there are so limited a number of major sentence units, it is difficult to make up probable distracters of word-order items. Compare the two items below. The first had a delta of 10.5; the second, of almost the same length, had a delta of 16. Differences in the probability of the distracters is clearly an important factor in explaining the large difficulty gap.

Example: Checked is by the boss the correspondence of the girls.
The boss is checked by the girls of correspondence.
The girls the correspondence is checked of by the boss.
The correspondence of the girls is checked by the boss.
The correspondence of the girls is by the boss checked.

Example: A new laboratory will be completely added to the campus facilities.
A completely new laboratory will be added to the campus facilities.
A new laboratory will be added completely to the campus facilities.
A new laboratory will be added to the campus facilities completely.
A new completely laboratory will be added to the campus facilities.

An extreme example of improbable distracters is the following set of responses to an item with a 9.8 delta:

should to go
should go
ought go
should goes
ought goes

Since all of the distracters are structurally impossible, the lead for this item was superfluous.

Similarity between item format and pattern-practice teaching materials was one of the factors suggested above to explain the high percentage of correct responses to word-order items. The same factor is probably operative in the large number of correct responses to question-word and echo-verb items, which patterned as follows:

| | Group I | II | III | IV |
|---|---|---|---|---|
| Question-word items | 13 | 0 | 1 | 0 |
| Echo-verb items | 8 | 5 | 4 | 2 |

The low deltas on the question-word items are not suprising, since content as well as form should make them easy: question words have a high frequency of occurrence, are for the most part morphologically invariable, closely parallel their Spanish equivalents in usage and meaning and are regularly tested through the question-answer pattern in which they were learned.

Echo verb structures, on the other hand, are linguistically complex and have no close parallel in Spanish. The predominantly low delta pattern of these items must be atrributable to successful teaching efforts in this area and to the similarity between commonly used practice drills and the forms of test items.

Under formal characteristics we can also consider linguistic factors which are not relevant to the material being tested. For example, to complete the lead *I went to the shoeshine parlor and* _____ , 312 students chose the correct response, ***had my shoes shined***, while 282 chose the distractor ***shine my shoes***, giving the item a rather high delta of 15.3. Part of the difficulty may have been that students were unfamiliar with the complement structure in the key, but this would not explain why so many chose ***shine my shoes***; most simple past tense items had much lower deltas. What probably does help to explain the pull of ***shine my shoes*** in a clear past-tense context is that the dental "marker" in a interconsonant position is practically never pronounced by Spanish speakers and is often omitted from informal native English speech as well. A Spanish speaker who has learned English phonemically, then, but who is weak on the graphic sytem, would tend to "see" ***shine my shoes*** as past, particularly in the presence of other past-tense markers.

In another item, with 14.9 delta, 366 students incorrectly chose the distracter ***send***, while only 232 chose ***sent***, the obvious key. It seems likely that a large number of students simply did not distinguish between the final graphemes of *send* and *sent*, since these represent phonemes which are often omitted or substituted for one another in speaking.

Without the support of data from specially designed experiments, of course, these observations can be offered only as educated guesses. They suggest how difficult it is to pinpoint the precise level of difficulty of a given item or to describe all probable causes of that difficulty.

Two other formal differences were studied, but no significance was established for them. The first was number of words in item and key. Group I items averaged 8.2 words each, as compared to an 8.86 word average for the Group IV items. The difference is meaningless in the context of all the other variables of form and content.

A small number of items required the student to pick the *wrong* answer; in other words, the four distracters were all acceptable in context, and the key was unacceptable. A few items of this sort were found in all groups, with the largest number in Group II. They seemed to have about the same delta patterns as their choose-the-right answer counterparts, but it is always possible that, had they been present in larger numbers, significantly different patterns would have become evident.

Among the 400 items studied, there were two mistaken keys, one double key and two misleading leads. This is a remarkably good score. Obviously ESLAT scores are not significantly affected by carelessness or mechanical error.

### 3. *Linguistic Factors in Delta Scores*

Apart from the form of an item and the success of the distracters, there are characteristics in the particular material being tested which make it more or less difficult for those who take ESLAT. The number of such characteristics is undoubtedly quite large, and their complete evaluation would require an individual analysis of each item, a procedure whose usefulness is questionable, however interesting it might be. The procedure here will be to define and illustrate the most important linguistic factors affecting item difficulty and then, in succeeding sections of this report, to show how these factors operate on student responses to certain areas of the grammar.

The fact that there are linguistic factors which affect the difficulty of items should not mislead anyone into thinking that certain areas of grammar are intrinsically more difficult than others. If that were true, then native speakers of English would have the same difficulties as non-native speakers, and such is not the case, although some difficulties are common to both groups. Far more so than with native speakers, the priorities of the English programs in Puerto Rico determine the students' differing levels of success in different areas of grammar; unlike a native English speaker in an English-speaking environment, the Spanish-speaking student in Puerto Rico depends to a very large extent upon classroom instruction for his knowledge of English. ESLAT scores therefore reflect the emphases of local English programs.

The first two linguistic factors affecting difficulty must be viewed in close relationship to one another. *Frequency* refers to the number of times a feature of the language occurs in common usage. The past tense of regular verbs, the plural grapheme for most nouns, the forms of *be* and the personal pronouns all have a high frequency. The present subjunctive, the verb form *hanged*, the use of prepositions before interrogative and relative pronouns and the plural of the word *gymnasium* are all low frequency features. Other things being equal (a condition that will remain understood through the rest of this discussion), ESLAT items testing high-frequency features were very easy and those testing low-frequency features were hard. Exceptions to this general tendency will be discussed in the following sections.

The *range* of applicability of a feature refers to the number of different enviroments it occurs in. For example, the graphemes *(s)* and *(ed)* to mark plurality in nouns and past tense in verbs have a very broad range because they occur on almost all nouns and verbs in English. The personal pronouns and the forms of *be*, on the other hand, are unique and thus have an extremely narrow range. Broad range items tended to be easier than those of narrow range, but very high-frequency features were found to be easy despite their narrow range: the personal pronouns and the forms of *be* are good examples. When frequency was not so high, range apparently became more important. For example, the use of *were* as a so-called "past subjunctive" form in sentence like *I wish I were sure* and *If he were here, we could leave* has a fairly high frequency, yet it was very difficult on the ESLAT results. The most probable reason for the difficulty is its narrow range: nowhere else in the language does *were* take a singular subject or refer to present time. Despite the narrow range of this feature, its frequency entitles it ot more attention than it seems to have been receiving in our ESL programs.

*Variability* is another factor influencing difficulty. It was suggested above that one reason for the large number of word-order items answered correctly was the fact that word-order patterns in English sentences is limited: it has low variability. On the other hand, the English verb phrase has a large number of possible combinations of verb forms, and this is one of the reasons for the difficulty of this area of grammar. Special problems seem to arise when a second language has a higher variety for a given feature than the student's native language. For example, when verbal structures function as complements in Spanish, they appear in the infinitive form; in English, however, they may appear either in the infinitive or in the *-ing* form *(He likes to swim* but *He enjoys swimming)*. As might be expected, the *ing* complements were found mostly among the hardest items; six out of eight complement items in Group IV were of this type.

The difficulty factor which operates most independently and through the greatest number of item types is *interference*, which occurs when a pattern from the first language prevents the correct pattern selection—or promotes the selection of an incorrect pattern—in the second language. Interference is a very powerful factor, operating on features of very high frequency and causing large differences between the delta scores of otherwise similar items. As an illustration of what has been said, let us look at several items testing *other* in different contexts. *Other* is a high frequency feature and behaves very similarly

to *otro* in Spanish. One important difference, however, is that when used as a modifier *other* cannot take, as *otro* does, a plural marker to indicate a number agreement with a plural head. When it is used as a modifier with a plural marker, interference is the obvious explanation. Now examine the following items. The number before each item is the delta score. The number following each alternative is the number of students who chose that alternative. In parentheses are the correct Spanish equivalents for the keys. Notice that for the high-delta item, the correct Spanish equivalent resembles the preferred distracter.

15.9 John knows how to drive a car. However he realizes that he has to know _____ things too.

| | | |
|---|---|---|
| other's | 18 | |
| others | 67 | |
| other | 85 | *(otras)* |
| another | 54 | |
| another's | 20 | |

11.00 A few of the students in this class are lazy, but ____ are all responsible.

| | | |
|---|---|---|
| other | | |
| other | 79 | |
| another | 76 | |
| the others | 599 | *(los otros)* |
| the other | 73 | |
| any other | 60 | |

9.9 William has two dogs. One is called Tiger. _____ is called Leopard.

| | | |
|---|---|---|
| The others | 21 | |
| The other | 717 | *(El otro)* |
| Another | 93 | |
| The others dog | 69 | |
| Other dog | 39 | |

The last two of these items are examples of *reinforcement*, in which the choice of the correct pattern in the second language is facilitated by a similar pattern in the native language. Because many of the items studied for this report are still in active use, no additional examples of interference will be reproduced here. Suffice is to say that at least 20 per cent of the items in Group IV were difficult because of interference, as determined by an examination of the distracters. No intereference patterns were found in the Group I items. Close comparison of Groups I and IV turned up five additional pairs of items with delta deifferences like those just cited: each pair tested material of comparable frequency, range, variability and subject matter, yet the members of each pair were separated by between five and eight delta points. All of the hard items had intereference difficulties; none of the easy items did. So despite the considerable amount of work that has been done in English-Spanish contrastive analysis, interference is still a major stumbling block for Spanish speakers trying to master English as a second language.

Educators have cautioned against an overuse of interference on ESL tests. They argue that contrasts likely to produce interference constitute a relatively small number of items in the whole of English grammar; if a disproportionate number of such items appear on a test, the test results will not accurately reflect the students' true proficiency. The caution is valid, certainly. Nevertheless, interference creates problems in many features—it is impossible to say exactly how many—of high frequency and broad range, and any

attempt to limit the number of such features would also lead to distorted test findings. The ESLAT materials were well selected in the sense that nearly all of the interference items tested basic language structures of broad range and high frequency. The caution mentioned above is directed against the excessive use of *ad hoc* idiom or vocabulary items which may be difficult in themselves but which reveal little about the students' control over the language as a whole.

One identifying characteristic of interference items is the fact that, although they are very hard for learners of English as a second language, they present no problem for native speakers. *Formal usage* items, on the other hand, are difficult for native and non-native speakers alike. They typically test features of low frequency and narrow range occuring under clearly prescribed circumstances. Although they make up an extremely small proportion of normal English usage, they consume large stretches of English class time under the designation of "correct grammar." Examples of formal usage include, for example, the selection of *whom* instead of *who* to mark the "objective case," the use of the subject forms of the personal pronouns in complement positions after *be* and certain conventions of subject-verb agreement. One of the reasons why these features of the language are especially difficult and have to be learned at school is that, in addition to being low frequency and narrow range, many of them run counter to what Edward Sapir (*Language*, 1921) called the "drift" of the language; some are chance survivals from earlier stages in the history of the language, while other are impositions from foreign grammatical models or from conceptual approaches to grammar which have been discredited. Although the ESLAT materials did not contain an overwhelming number of formal usage-items, the use of any of these items is open to question. For this reason a separate section of this report will be devoted to them below.

The object of this section has been to show that ESLAT items vary in difficulty according to the interplay of five major factors—frequency, range, variability, interference and formality of usage—and to suggest that of some of these factors are better norms that others. The next section will chart the language areas tested by 400 items divided into four levels of difficulty. After that, significant correlations between certain language areas and difficulty levels will be discussed.

4. *Language Features at Four Levels of Difficulty*

| | Group I | Group II | Group III | Group IV |
|---|---|---|---|---|
| Word order | 33 | 15 | 7 | 4 |
| Main verb phrases | 10 | 28 | 40 | 34 |
| Echo verb forms | 8 | 5 | 4 | 2 |
| Preposition choice | 8 | 11 | 10 | 2 |
| Context sensitive words | 3 | 3 | 4 | 11 |
| Correct pronoun form | 12 | 11 | 8 | 3 |
| Question-word responses | 13 | — | 1 | — |
| Adverbs in negative contexts | 2 | 5 | — | 6 |
| Pre-head noun modifiers | 4 | 4 | 5 | 6 |
| Verbal complements | 4 | 4 | 3 | 8 |
| Other adverbs | 1 | 1 | — | 2 |
| Comparison structures | 2 | 6 | 5 | 7 |
| Agreement problems | — | 1 | 4 | 8 |
| Possessive problems | — | — | — | 7 |
| Pronunciation | — | — | 1 | — |

The selection of features for a division like this is always somewhat arbitrary, a compromise among numerous considerations. For example, why split echo verbs from main verbs? In that case the decision was easy: first, echo verbs are a limited form of response to preceding verb forms, whereas main verbs are a response to a total syntactic and semantic context; second, main verbs are most numerous in the two highest difficulty groups, while echo verbs occur in inverse ratio to the level of difficulty. Difficulty-level clusters were responsible for a number of individual groupings, including the "possessive problems" which were found only in Group IV. On the other hand, additional subdivision might have been made. Some of the main verb items tested the forms of verbs and verb phrases (like the *should go* item referred to in Section 2) while others test a choice among equally acceptable forms only one of which fits a given context. But there is always the risk of losing a clear picture of larger numerical patterns in a welter of subdivisions, so these were generally avoided unless they seemed likely to yield significant information. In the case of the main verb phrases, which were more numerous than any other language feature, too many different factors had to be considered to make subdivision a practical method of analysis. They will be discussed in a separate part of the report.

Finally, many of the items tested more than one grammatical feature. For instance, selection of a verb phrase might be conditioned by time signals in the lead, by the type of complement present, by a negation structure, by agreement requirements with the subject and so forth. Wherever this problem existed, assignment to a particular grammatical category was decided upon by a study of the distracters, the strongest distracter being used as the clue to what the item was principally testing. The distracters were also the main source of information about the causes of difficulty, as illustrated in the discussion of interference in the previous section.

5. *Main verb phrases*

The transformational grammarians have shown that the rules underlying verb-phrase formation in English are few in number and easily applied. The number of surface level structures that can be produced by these rules, however, is extremely diverse. To see this one has only a compare a transformational description of the verb phrase, which can be expressed in a simple, elegant formula, with the complicated surface-level description in a structural grammar like W. Nelson Francis's *The Structure of American English*. The native speaker of English produces verb phrases by applying the simple underlying rules, but the learner of English as a second language is faced with the high variability of surface level formations. This high variability of verb phrases undoubtedly helps to explain why there were so many of them among the ESLAT items and why there were nearly twice as many at the higher levels of difficulty than at the lower.

Frequency also explains a great deal about the difficulty distribution of the verb phrases. In terms of usage, the most frequent verb phrases in English are those consisting of one or two words: ***goes, was going, may go, etc.*** The longest possible verb phrase is five words long ***(He must have been being questioned for hours)*** and has an extremely low frequency; for most speakers it probably has zero frequency. Indeed, most English speech and writing seldom uses more than three words in a verb phrase, and the number of three-word units is far lower than those of one or two words. Now among the verb phrases in Group I and II none was longer than two words, and the majority in Group I were single-word past or present forms. Thus the items showed a close relation between length, frequency and successful response.

Length was not the only criterion of frequency to be considered. Group IV contained three items testing the present subjunctive, a form that has been dying out since the beginning of the modern English period and that is now limited to a single environment of very low frequency. One of the present-subjunctive items had the highest delta (19.7)

of all the items examined; the item was *I thought it advisable that she be chosen for the position.* A sentence like this is so improbable and the item gives so little information that it should probably be eliminated from the ESLAT files. The usage, in addition to being very low frequency and narrow range, is very formal; most speakers under most circumstances would say something like *I thought she ought to be chosen for the position.*

The situation with the past subjunctive is different. The range of the past subjunctive is very narrow: it occurs mostly in subordinate clauses after *wish (I wish I were rich)* and in contrary-to-fact conditions *(If I were you, . . .).* The form *were* has an even narrower range than other past subjunctive forms, since nowhere else does it occur with a singular subject. Resistance to the use of *were* with a singular subject is strong even among native speakers, who sometimes say *I wish it was time to leave already* and *If I was you, I'd quit.* On the ESLAT items, however, the strongest distracters were those with verb forms which normally refer to present or future item. This indicates that many students do not recognize that "past tense" forms in English can serve a subjunctive function that has nothing to do with past time (Interestingly enough, equivalent verb forms in Spanish —*supiera, fuéramos, estuvieran*— use the same base as past tense forms.) Group IV had 13 past subjunctives referring to present or past time *(If he were here, If he had been here).* Group III had six past subjunctives, all of them referring to past time. The past-time verb structure, *had* plus the past participle, produces no agreement anomalies and it clearly marks the time referred to, so one would expect it to be somewhat easier than the present-time forms. Nevertheless, no subjunctive usages at all appeared in Groups I and II, indicating that such usages are very difficult for the majority of those who take ESLAT.

Another set of tenses which proved difficult were the perfects and perfect progressives *(have/had lived, been living).* This was unexpected since these tenses have broad range and high frequency and are structurally identical to the equivalent forms in Spanish. Nevertheless, they appeared mostly in the high delta groups:

| Group | I | II | III | IV |
|---|---|---|---|---|
| Perfect tense items | — | 3 | 10 | 10 |

Analysis of the distracters indicated that the difficulty was caused principally, by two factors, both involving the choice of a tense appropriate for a given environment. First, interference was obvious in items like *They have been living here since 1970*, where the distracter *are living* would predictably draw a large number of responses on the model of the Spanish *están viviendo.* Second, the choice between perfect and simple past is often doubtful. For example, for the item *Have the Smiths had time to get to Ponce?, Did the Smith have* was a powerful distracter. It would have been a good distracter for native speakers as well: *They've just left* and *They just left* are equally probably responses to a question like *Are they there?* in contemporary English. The slipperiness of the perfect-past distinction even tripped up one of CCEB's item writers: *have been* was given as the key for *There____ several calls for you while you were gone*, whereas *were* would be the appropriate choice in the presence of the subordinate clause marking the past.

The subjunctives and the perfect tenses were the only verb phrases that showed consistently strong difficulty patterns. In most cases the appropriateness of a given verb phrase to a given context was being tested, rather than the structure of the verb phrase; but even in Group IV there were a few items testing irregular verb forms. One- and two-word indicative phrases other than the perfects were most common in the easier groups.

### 6. *More on Interference*

The ESLAT items suggest that interference can operate in any area of grammar. Even word order, one of the easiest areas, can be difficult in the presence of interference

problems. The following item asking for the correct choice of a sentence had a delta of 17.2. The chief distracter follows the Spanish word order for that type of subordinate clause, while the key is foreign to Spanish word order.

| | |
|---|---|
| You do remember when is the dance? | 84 |
| Do you remember when the dance is? | 252 |
| Do you remember when is the dance? | 846 |
| When the dance is you do remember? | 82 |
| When is the dance do you remember? | 279 |

The fact that almost any item can be made difficult if strong interference patterns are present among the distracters is the main reason why care and restraint must be exercised in the use of such distracters.

There are certain contrasts, however, in which interference is so common that it becomes a sort of built-in barrier for the Spanish speaker trying to learn English. This is most obvious in phonology, but quite marked in some morphological and syntactic patterns as well. For example, all items on the English double possessive *(a friend of hers)* were found in Groups III and IV, despite the very high frequency of the structure in English. The reason why interference operates so powerfully here is that the competing structure in Spanish, *un amigo de ella*, looks very similar, so the Spanish ESL speaker is strongly drawn toward *a friend of her*. It is interesting that all but one of the double possessives fell in group IV, and the one that fell in Group III was *a cousin of mine*: Spanish does not have a construction *un primo de mí*, so the interference is apparently less.

Another feature troubled by interference is the *-ing* verbal complement, as in *He enjoys swimming*. The interference here stems from a difference in variability: where Spanish uses the infinitive as verbal complement, English uses either the infinitive or the *-ing* form depending on the main verb. There is no morphological equivalent to the *-ing* complement in Spanish, so the Spanish speaker is drawn to use the infinitive—*He enjoys to swim*—for all verbal complements in English. Of the seven ESLAT items requiring the selection of an *-ing* complement, six were in Group IV and one in Group III. Among the easier groups, all the verbal complements were infinitives *(Carmen wants to study at the University next year)*, a pattern with strong reinforcement from Spanish. Also, the *-ing* complement has a narrower range than the inifinitive complement in English, and with verbs that allow either type of complement *(like, continue)* the infinitive is chosen more often.

Group IV contained a number of interference items of very narrow range: *people are, news is, some furniture* and several others whose difficulty could easily have been predicted by anyone who has ever taught ESL to native speakers of Spanish. There is a well established list of such items which can be counted upon to produce high deltas; it's like pushing a button to get a wrong response. Despite the high frequency of some of these expressions, however, many of them are *sui generis*, so their use in test items is rather sterile insofar as giving information about a student's total command of English is concerned. Suppose a student knows that *people* is a plural noun? What does this isolated bit of information tell us? Granted that a student who recognizes that *people* is plural is likely to know a great deal more about English, from a linguistic point of view it would be preferable to test features of equal difficulty but broader range.

The question of range is not limited to interference patterns. One difficult item tested the selection of the word *shrug* in *He . . .gave a great shrug of his shoulders.* Again, this item may be of some value in signaling students with considerable experience in English. But it is far less revealing than the item *He did the work well*, which was also from Group IV, but which tested a context-sensitive word of far broader range and higher frequency. It also has a 23 points higher reliability than the *shrug* item.

### 7. *Formal Usage and Other Problems*

The difficulties just discussed are difficulties for the learner of English as a second language, not for native speakers. Among the ESLAT items there were also some which would have proven fairly hard for a group of native speakers at the high school level. The first of this type test pronouns in contexts where correct usage requires that the regular morphological-syntactic correspondences be reversed, as in the use of subject forms in complement positions and object forms in pre-verb position. Both of these anomalies are present in the item ***Was it really she whom you saw last night?***, a question whose real-life probability is so low that it sounds perverse! This is a good example of the sort of usage, described in Section 3 above, which stands against the drift of the language, like an unyielding rock in a living stream. But the real objection to the item is that it tests a pattern which is not really used outside of school grammars. The pattern which *does* occur and which ought to be tested is ***Was she the one you saw last night?*** As a matter of policy ESLAT should admit only high probability items—items, that is, with a high frequency-range combination.

A second group of special-difficulty items were those which tested logical or structural problems in agreement. One such problem can be observed in ***Our chief problem is the students***, where the verb form is torn between the singular noun before and the plural noun after. Interference also affected the results on this item, since the equivalent Spanish sentence should require the plural verb form.

A more common language-logic problem centered on words like ***each*** and ***every***, which are singular but which frequently serve as head words for constructions which are necessarily plural in referent. Thus, ***each of the professors has his own car*** takes a singular verb because ***Each*** is seen as the subject, despite the fact that ***each of the professors*** has the same referent as ***all of the professors***. It is interesting that when ***a lot*** or ***all*** serves as the head of this type of construction, the word following the preposition determines the number of the entire unit: ***a lot of money*** is singular and ***a lot of people*** plural. As there are other areas of the language where a whole headed unit rather than the head word enters into grammatical relationships ***(the man on the moon's nose)***, it is understandable that ***each*** and ***every*** should lead to divided usage. The division, however, is still clearly along the line separating standard from non-standard.

The third type of agreement problem encountered in the difficult groups was rather contrived and mechanical: it consisted of a subject separated from the verb by a phrase containing a noun of different number. The item ***All the rooms, except one, are air conditioned*** fell into Group IV because of the distraction from ***except one*** in an otherwise simple case of agreement.

The final area of special difficulty which should be mentioned is the phrasal genitive, the addition of the genitive marker to a whole phrase rather than the particular word to which it logically belongs, as in ***sister-in-law's***. The phrasal genitive is no problem for native speakers but for native Spanish speakers, whose language has no inflected genitive of any sort, the problems are formidable. (Among English speakers divided usage is starting to become apparent in applying the plural marker to a whole phrase: ***sister-in-laws*** is now heard frequently beside ***sisters-in-law***.)

Except for the phrasal genitive, which has developed a broad range in modern English, the appropriateness of the materials discussed in this section should be re-examined for future versions of ESLAT. The fact that the items can be relied on for high deltas may make them useful in prediction and test design; it has already been pointed out, however, that difficult items can also be written to test features of high frequency and broad range, features which are unquestionably at the living core of contemporary Enlgish. Most of the features discussed above show some degree of divided usage among native speakers of English; whether such material should be included on a test designed for ESL students is a question to be thought about and perhaps worked into policy.

The reliability scores for most of the items discussed in this section were below .40, well below average though not below minimum acceptability.

8. *Some Observations on Reading Comprehension Items*

In addition to the structure items discussed in the first part of this report, 10 reading selections were examined, each accompanied by five questions. Each question had its own delta, and the average of the five deltas for each selecton was given as the delta for the unit as a whole. The following list gives the deltas for each of the ten units examined; the figures in parentheses show the spread of the deltas for the individual items in the unit.

15.54 (13.0 - 16.8)
15.40 (14.3 - 16.6)
15.34 (13.6 - 16.7)
15.20 (12.9 - 17.0)
15.16 (12.3 - 18.8)
15.08 (11.7 - 17.0)
14.48 (12.9 - 15.5)
14.44 (13.7 - 15.7)
14.10 (12.9 - 14.6)
14.00 ( 9.9 - 15.8)

The most striking characteristic of this list is the similarity of the unit deltas, which vary by the narrow distance of from 14.00 to 15.54. And except for the 9.0 and 16.8 figures, there is also a remarkable similarity in the delta spreads for the individual items.

These figures might lead one to suppose that the 10 reading selections were about equally difficult. Nothing, however, could be farther from the truth. The range of difficulty was very great, and there was no relationship pattern between the difficulty of the selections and their delta scores. For example, the 15.54 delta unit had as its reading selection a simple newspaper report on a meeting of a psychology association. The unit with the 15.34 delta, on the other hand, was a highly technical treatment of the medical uses of radioisotope rays.

It was assumed, then, that if the delta score did not reflect the difficulty of the reading selection, it had to reflect the difficulty of the items or of the items in relationship to the selection. In other words, the more difficult selections would be accompanied by easier items, and vice versa. For this assumption to be proven, the characteristics of hard items would have to be recognized and distinguished from the characteristics of easy items. The recognition of at least some of these characteristics emerged rather quickly: an item was easy if the key repeated the same words that contained the required information in the reading selection, especially if the words were in close juxtaposition. This can be seen in the following low-delta items compared with the contextual clues necessary to answer them.

| Delta | Item | Contextual clue |
|---|---|---|
| 13.4 | The reading states that radium is used in the treatment of cancer because it *destroys tissue.* | The use of the powerful tissue destroying rays from radium is established in the treatment of cancers. |
| 13.0 | *. . .a pill may curb aggressive behavior.* | . . .pills or other treatments to curb aggressive behavior. . . . |

When the information in the context was paraphrased in the key, even in a relatively direct way, the delta was significantly higher.

| | | |
|---|---|---|
| 16.2 | According to the preacher he never got hurt because he was *free of sin.* | "The preacher said. . ., 'The innocent walk with God—that's why he don't get hurt." |
| 16.6 | . . .Mr. Whipple talked all the time about her son probably because *she got comfort this way.* | . . .she had to talk about him first, before she could get on to anything else. It seemed to ease her mind. |

When a word or word used prominently in the selection happend not to be the key, it usually proved to be a strong distracter. For example, the last choice *(all of the above)* in the following item was the obvious key, yet *embarassed* was chosen more than twice as often:

16.0 The husband was

| | |
|---|---|
| angry | 21 |
| embarrassed | 284 |
| stupid | 11 |
| unkind | 53 |
| all of the above | 114 |

The reason so many people chose embarrassed was that the selection contained the statement *he was hotly embarrassed and indignant at his wife for embarrassing him.* The only word among the distracters that appeared in the selection was *unkind.*

Items were also relatively easy when their keys contained information implicit through the whole reading selection. Thus, the item with the lowest delta (9.9) was *The wife wept because her husband was ungrateful.* The ingratitutde of the husband in question was apparent in almost every line of the reading selection. Sometimes a bit of cultural background probably played its part in leading students to the key, as in the (delta 12.3) item *The new way of life mentioned by Pasteur is probably a life of study and dedication to humanity.*

There was no indication that the item writers in reading comprehension were testing for recognition of specific grammatical structures. Some of the items required that inferences be drawn from the given data. But vocabulary recognition was easily the most frequent determinant of difficulty. The highest delta item among the reading group was based on simple word equivalence: *The expression barren skepticism in the second paragraph means fruitless doubts.* This item could obviously have been answered correctly by persons who had never seen the reading selection.

However effective the reading items for ESLAT may be statistically, the preceding discussion brings up some serious questions about the skills they measure. First, if the delta scores for a given set of items do not indicate the difficulty of the reading selection on which they are based, of what use are they? And is their statistical significance not limited by the narrow range in which most of the delta scores cluster? Second, the apparent absence of a design to test recognition of specific linguistic structures raises the question of how accurately the items measure overall strengths and weaknesses. Third, when the correct response to an item can be made through an almost mechanical association between identical vocabulary items in the reading selection and in the key, a very low level of reading proficiency is required; some degree of indirection, the use of paraphrase at least, should be involved.

These observations are based upon examination of a small sample, so they are not intended as a general criticism of the reading items. They are offered as indications of possible weaknesses to be kept in mind with a view to possible improvements.

9. *Textbooks and Test Scores*

An attempt was made to relate the ESLAT results to the emphases in textbooks used in the public high schools, which educate the great majority of students in Puerto Rico. The English Section of the Department of Public Instruction kindly provided copies of the following texts for this purpose: Books, 4, 5 and 6 of the **Lado English Series**, with accompanying workbooks; Book 4 of the **Let's Learn English** series; and Books 1, 2 and 3 of the **English for Today** series. The Lado series is the basic source of instructional material, being used by about 75 percent of the students in the English program. The others are used by limited experimental groups.

It was not possible, of course, to make any sort of correlation between item success and textbook used; nor was it possible to learn the number of private school students included in the numbers that responded to a given item. All that could be assumed was that most of the students responding to a given item had used the Lado series for the previous three or four years. Under these conditions very little could be proven about the merits of any of the instructional materials. It was decided, then, simply to examine each of the series to see if it covered the grammatical areas found to be more difficult on ESLAT. The areas selected were the perfect tenses, contrary-to-fact conditions, noun clauses after *wish*, verbal complements in *-ing* and the double possessive.

The **Lado English Series** fared the worst in this comparison. It covered only the *-ing* complement, contrary-to-fact conditions in present time, and the perfect tenses. Moreover, the treatment given to the perfect tenses was inadequeate: explanation and exercises together covered about a page and a half in Book six, and no attempt was made either to relate the tenses to specific contextual clues or to distinguish between the use of the perfect tenses and other tenses which are sometimes incorrectly used in their place. The treatment of the contrary-to-fact conditionals was also, from a pedagogic point of view, insufficient.

This evaluation is based on an examination of only the last three books of the six-books series. Undoubtedly the double possessive is covered in one of the earlier books. But one might reasonably expect that all areas of exceptional difficulty should be at least reviewed during the final three books. The high difficulty patterns of items relating to these areas indicate clearly that they require greater instructional emphasis.

Book 4 of the **Let's Learn English** series is the second volume of the intermediate level. In Puerto Rico it is being used in an experimental program in the last year of high school. It covers contrary-to-fact conditions, noun clauses after *wish* and verbal complements in *-ing*. It even covers the present subjunctive, which neither of the other series mentions. The material is very well presented and organized, and the exercises are intelligently planned to cover difficulties that the explanations do not fully resolve. The workbook to accompany this text was not available for examination.

Books One, Two and Three of **English for Today** start at the very earliest stages of ESL and present with great thoroughness a limited but very well chosen sample of the language. The extensive treatment of the perfect tenses in Books Two and Three is superb, with a detailed presentation of contextual clues and with pages and pages of practical exercises. The series also contains contrary-to-fact conditions in present time, the past subjunctive in noun clauses after *wish*, and the double possessive. Excellent workbooks supplement the generous exercises in the text.

Like most current ESL series published commercialy on a large scale, these three are aimed at as broad a second-language market as possible, so they cannot afford to give

special attention to the linguistic contrasts which cause interference for the Spanish-speaking ESL student. Ideally, then they should be supplemented with locally prepared materials concentrating on areas of persistent interference.

10. *Conclusions and Recommendations*

In Section (1) the hope was expressed that this report would lead to further experimentation and research on English teaching in Puerto Rico. One obvious possibility might be a joint project of the College Board and the Department of Public Instruction to use ESLAT as a device for evaluating the relative effectiveness of different sets of teaching materials. A project of this sort would take some years but should pose no great administrative problem and would be immensely useful. A simpler project might be a comparison of ESLAT scores obtained by students from different school systems; the information could lead to a fruitful comparison of materials, techniques and expectations in the different systems. Also, a comparison between the performance of native English speakers and native Spanish speakers might turn up a great deal of new information about the problems of second-language learning.

The use of ESLAT for projects of this sort would probably require that future versions of the test contain a more careful balance of linguistic materials. Some imbalances in the present collection of ESLAT items can be deduced from the language features chart given above. More than 25 percent of the items tested main verb phrases, and many items were repetitious, testing the same structure with different verb heads. The number of word-order items also seems excessive, especially the heavy concentration in Group I. Conversely, there relatively few items testing the meanings of modals, indirect discourse, *there*, *was/were*, transformational relationships between superficially dissimilar structures, verbal modifiers and other important features. It seemed as though the items writers had begun with a fixed and limited idea of what things ought to be tested, an idea based on their past experience with tests and teaching materials. They might find it more interesting to begin with a small corpus—a few pages from a contemporary novel or play—and try testing for all the language features they find there.

Basically, however, ESLAT is an excellent testing instrument. It has already made available a great amount of information which the English teaching profession can and should make use of. The most important single observation made in this report is that Spanish-English interference problems are still a major source of difficulty for ESL students, even though numerous studies, including those by researchers in Puerto Rican universities, pinpoint such problems with clarity and precision. (See, for example, *Readings in Spanish-English Contrastive Linguistics*, edited by Rose Nash and published by the Inter American University Press.) These studies should be used as the basis for supplementary materials to meet the specific needs of Puerto Rican students. One area which obviously needs more attention is the use of the perfect tenses, particularly the contextual clues which require their selection. More emphasis should also be placed, in local English programs, on double possessives, *-ing* complements, structures involving negative adverbs, common context-sensitive words and noun modifiers.

Two very positive conclusions emerge from this report. The first—based upon the large proportion of successful responses to items on word order, question words, pronoun selection and echo verb forms—is that good teaching programs will produce good results in the great majority of students learning English as a second language. The second is that available tests such as ESLAT can be used to discover weaknesses in existing programs and to point the way toward more effective programs in the future.